SEQ ID 1

```
LOCUS       HDU53215                2384 bp    DNA     linear   BCT 15-MAY-1997
DEFINITION  Haemophilus ducreyi cytolethal distending toxin protein A (cdtA),
            protein B (cdtB) and protein C (cdtC) genes, complete cds.
ACCESSION   U53215
VERSION     U53215.1  GI:2102681
KEYWORDS    .
SOURCE      Haemophilus ducreyi
  ORGANISM  Haemophilus ducreyi
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 2384)
  AUTHORS   Cope,L.D., Lumbley,S., Latimer,J.L., Klesney-Tait,J., Stevens,M.K.,
            Johnson,L.S., Purven,M., Munson,R.S. Jr., Lagergard,T., Radolf,J.D.
            and Hansen,E.J.
  TITLE     A diffusible cytotoxin of Haemophilus ducreyi
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 94 (8), 4056-4061 (1997)
  MEDLINE   97268696
   PUBMED   9108104
REFERENCE   2  (bases 1 to 2384)
  AUTHORS   Hansen,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-1996) Microbiology, UT Southwestern Medical
            Center, 6000 Harry Hines Blvd NA6.200, Dallas, TX 75235-9048, USA
FEATURES             Location/Qualifiers
     source          1..2384
                     /organism="Haemophilus ducreyi"
                     /strain="35000"
                     /db_xref="taxon:730"
     gene            152..823
                     /gene="cdtA"
     CDS             152..823
                     /gene="cdtA"
                     /codon_start=1
                     /transl_table=11
                     /product="cytolethal distending toxin protein A"
                     /protein_id="AAB57725.1"
                     /db_xref="GI:2102682"
                     /translation="MKKFLPSLLLMGSVACSSNQRMNDYSQPESQSDLAPKSSTIQPQ
                     PQPLLSKTPSMSLNLLSSSGPNRQVLPSEPSNFMTLMGQNGALLTVWALAKRNWLWAY
                     PNIYSQDFGNIRNWKMEPGKHREYFRFVNQSLGTCVEAYGNGLIHDICSLDKLAQEFE
                     LLPTDSGAVVIKSVSQGRCVTYNPVSTTFYSTVTLSVCDGATEPSRDQTWYLAPPVLE
                     ATAVN"
     gene            838..1689
                     /gene="cdtB"
     CDS             838..1689
                     /gene="cdtB"
                     /codon_start=1
                     /transl_table=11
                     /product="cytolethal distending toxin protein B"
                     /protein_id="AAB57726.1"
                     /db_xref="GI:2102683"
                     /translation="MQWVKQLSVVFCVMLFSFSSYANLSDFKVATWNLQGSSAVNESK
                     WNINVRQLLSGEQGADILMVQEAGSLPSSAVRTSRVIQHGGTPIEEYTWNLGTRSRPN
                     MVYIYYSRLDVGANRVNLAIVSRRQADEAFIVHSDSSVLQSRPAVGIRIGTDVFFTVH
                     ALATGGSDAVSLIRNIFTTFNSSSSPPERRVYSWMVVGDFNRAPANLEVALRQEPAVS
                     ENTIIIAPTEPTHRSGNILDYAILHDAHLPRREQARERIGASLMLNQLRSQITSDHFP
                     VSFVRDR"
     gene            1700..2260
                     /gene="cdtC"
     CDS             1700..2260
```

```
                    /gene="cdtC"
                    /codon_start=1
                    /transl_table=11
                    /product="cytolethal distending toxin protein C"
                    /protein_id="AAB57727.1"
                    /db_xref="GI:2102684"
                    /translation="MKKYLLSFLLIMILALASHAESNPDPTTYPDVELSPPPRISLRS
                    LLTAQPVKNDHYDSHNYLSTHWELIDYKGKEYEKLRDGGTLVQFKVVGAAKCFAFLGK
                    GTTDCKDTDHTVFNLIPTNTGAFLIKDALLGFCITSHDFDDLKLEPCGGSVSGRTFSL
                    AYQWGILPPFGPSKILIPPVRRNQGS"
BASE COUNT        714 a     416 c     494 g      760 t
ORIGIN
        1 agaaaagata ttgaacaggt tctaacataa agtataataa aggttcatat gctttttttac
       61 gatattatct ggtttggttt tatcaaaaaa aaggataaaa tgcgaagaac ttgtcctttt
      121 aaatttaagg atggatctaa ggagagatat aatgaaaaag tttttaccta gtcttttatt
      181 gatgggttca gtggcttgtt catcaaatca acgaatgaat gactattctc aacctgaatc
      241 tcaatctgat ttagcaccta aatcttcaac aatacaaccc caacctcaac ccctattatc
      301 aaaaacacct tcaatgtcac tgaatttgct atcttcatcc ggaccgaata gacaggtatt
      361 gccgtctgaa ccatcaaact ttatgacttt gatgggacaa aatggggcac tgttgactgt
      421 ctgggcgcta gcaaaacgca attggttatg ggcttatccc aatatatatt cgcaggactt
      481 tggaaatatt cgtaattgga agatggaacc cggtaaacac cgtgaatatt ttcgtttttgt
      541 taatcaatct ttaggtacat gtgttgaagc ttacggtaat ggtttaattc atgatatttg
      601 tagtctggac aaattagcac aagagtttga gttattacct actgatagtg gtgcggttgt
      661 cattaaaagt gtgtcacaag ggcgttgtgt cacttataat cctgtaagta caacatttta
      721 ttcaacagtt acattatcag tttgtgatgg cgcaacagaa ccatcacgtg atcaaacatg
      781 gtatctcgct ccccctgtat tagaagcaac agcggttaat taaactaagg agtttatatg
      841 caatgggtaa agcagttaag tgtggttttc tgtgtgatgt tatttagctt ttcaagttat
      901 gctaacttga gtgacttcaa agtagcaact tggaatctgc aaggttcttc agcagtaaat
      961 gaaagtaaat ggaatattaa tgtgcgccaa ttattatcgg gagaacaagg tgcagatatt
     1021 ttgatggtac aagaagcggg ttccttacca agttcggcag taagaacctc acgggtaatt
     1081 caacatgggg gaacgccaat tgaggaatat acttggaatt taggtactcg ttcccgccca
     1141 aatatggtct atatttatta ttctcgttta gatgttgggg caaaccgagt gaacttagct
     1201 atcgtgtcac gccgtcaagc cgatgaagct tttatcgtac attctgattc ttctgtgctt
     1261 caatctcgcc ctgcagtagg tatccgcatt ggtactgatg tattttttac agtgcatgct
     1321 ttggccacag gcggttctga tgcggtaagt ctgattcgta atatcttcac tacttttaac
     1381 tcatcatcat ccccaccgga aagacgagta tatagctgga tggttgttgg tgatttcaat
     1441 cgtgcgccgg ctaatctgga agttgcatta agacaggagc ccgcagtgag tgaaaataca
     1501 attattattg cgccaacaga accgactcat cgatctggta atattttaga ttatgcaatt
     1561 ttacatgatg cacatttacc acgtagagaa caggcccgtg aacgtatcgg tgcaagttta
     1621 atgttaaatc agttacgctc acaaattaca tccgatcatt ttcctgttag ttttgttcgt
     1681 gatcgctaag gaggatatta tgaaaaaata tttattgagc ttcttattaa tcatgatatt
     1741 ggctttggcg agtcatgcag aatcaaatcc tgatccgact acttatcctg atgtagagtt
     1801 atcacctcct ccacgtatta gcttgcgtag tttgcttacg gctcaaccag ttaaaaatga
     1861 tcattatgat tcacataatt acttgagtac acattgggaa ttaattgatt acaaaggaaa
     1921 agaatatgaa aaattacgtg acggtggtac gttagttcaa tttaaagtgg ttggtgcagc
     1981 aaaatgtttt gctttcctgg gcaaaggcac aactgattgt aaagatactg atcatactgt
     2041 gtttaacctt attccaacta atacgggcgc gtttttaatc aaagatgcac tattagggtt
     2101 ttgtataaca agccatgact ttgatgattt gaagcttgaa ccttgtggag gttcagtgag
     2161 tggtcgaacc tttttcgttgg cgtatcaatg ggggatatta cctcctttttg gaccaagtaa
     2221 aattttaata ccaccggtgc gaagaaatca gggtagctaa tgttttacat ataattgtat
     2281 ttcttcaaat caagatcctt agtggggcga agaaatataa tgtcattatt gtgcttatgt
     2341 taatgatcat gcaaaaatga gccaggcaga cgcagtaaga tcat
//
```

CQ 10 2

```
LOCUS       CJU51121                2600 bp    DNA     linear   BCT 28-AUG-1996
DEFINITION  Campylobacter jejuni cytolethal distending toxin (cdtA), (cdtB),
            (cdtC) genes, complete cds, and LctP gene, partial cds.
ACCESSION   U51121
VERSION     U51121.1  GI:1354273
KEYWORDS    .
SOURCE      Campylobacter jejuni
  ORGANISM  Campylobacter jejuni
            Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
            Campylobacter.
REFERENCE   1  (bases 1 to 2600)
  AUTHORS   Pickett,C.L., Pesci,E.C., Cottle,D.L., Russell,G., Erdem,A.N. and
            Zeytin,H.
  TITLE     Prevalence of cytolethal distending toxin production in
            Campylobacter jejuni and relatedness of Campylobacter sp. cdtB gene
  JOURNAL   Infect. Immun. 64 (6), 2070-2078 (1996)
  MEDLINE   96239019
   PUBMED   8675309
REFERENCE   2  (bases 1 to 2600)
  AUTHORS   Pickett,C.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-1996) Carol L. Pickett, Microbiology/Immunology,
            University of Kentucky, UKMC 800 Rose St., Lexington, KY 40536, USA
FEATURES             Location/Qualifiers
     source          1..2600
                     /organism="Campylobacter jejuni"
                     /strain="81-176"
                     /db_xref="taxon:197"
     gene            141..947
                     /gene="cdtA"
     CDS             141..947
                     /gene="cdtA"
                     /codon_start=1
                     /transl_table=11
                     /product="CdtA"
                     /protein_id="AAB06707.1"
                     /db_xref="GI:1354274"
                     /translation="MQKIIVFILCCFMTFFLYACSSKFENVNPLGRSFGEFEDTDPLK
                     LGLEPTFPTNQEIPSLISGADLVPITPITPPLTRTSNSANNNAANGINPRFKDEAFND
                     VLIFENRPAVSDFLTILGPSGAALTVWALAQGNWIWGYTLIDSKGFGDARVWQLLLYP
                     NDFAMIKNAKTNTCLNAYGNGIVHYPCDASNHAQMWKLIPMSNTAVQIKNLGNGKCIQ
                     APITNLYGDFHKVFKIFTVECAKKDNFDQQWFLTTPPFTAKPLYRQGEVR"
     gene            944..1741
                     /gene="cdtB"
     CDS             944..1741
                     /gene="cdtB"
                     /codon_start=1
                     /transl_table=11
                     /product="CdtB"
                     /protein_id="AAB06708.1"
                     /db_xref="GI:1354275"
                     /translation="MKKIICLFLSFNLAFANLENFNVGTWNLQGSSAATESKWSVSVR
                     QLVSGANPLDILMIQEAGTLPRTATPTGRHVQQGGTPIDEYEWNLGTLSRPDRVFIYY
                     SRVDVGANRVNLAIVSRMQAEEVIVLPPPTTVSRPIIGIRNGNDAFFNIHALANGGTD
                     VGAIITAVDAHFANMPQVNWMIAGDFNRDPSTITSTVDRELANRIRVVFPTSATQASG
                     GTLDYAITGNSNRQQTYTPPLLAAILMLASLRSHIVSDHFPVNFRKF"
     gene            1752..2321
                     /gene="cdtC"
     CDS             1752..2321
                     /gene="cdtC"
```

```
                    /codon_start=1
                    /transl_table=11
                    /product="CdtC"
                    /protein_id="AAB06709.1"
                    /db_xref="GI:1354276"
                    /translation="MKKIITLFFMFITLAFATPTGDLKDFTEMVSIRSLETGIFLSAF
                    RDTSKDPIDQNWNIKEIVLSDELKQKDKLADELPFGYVQFTNPKESDLCLAILEDGTF
                    GAKSCQDDLKDGKLETVFSIMPTTTSAVQIRSLVLESDECIVTFFNPNIPIQKRFGIA
                    PCTLDPIFFAEVNELMIITPPLTAATPLE"
     gene           2502..2600
                    /gene="LctP"
     CDS            2502..>2600
                    /gene="LctP"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="AAB06710.1"
                    /db_xref="GI:1511637"
                    /translation="MLTWQQIYDPFSNIWLSALVAFLPILCFLVCLV"
BASE COUNT        873 a      402 c      421 g       904 t
ORIGIN
        1 tgctaaaata taagtgttta agatacatat aaattctacc tttaaaaaca acaaaataaa
       61 acatttttaa aaagcggaaa attataatga aatttatgtt attattttct taaaaattta
      121 aatacatatc aaggttttta atgcaaaaaa ttatagtttt tattttatgt tgttttatga
      181 ctttttttct ttatgcatgt tcttctaaat ttgaaaatgt aaatcctttg gggcgttcat
      241 ttggagaatt tgaagatact gatcctttaa aactaggact tgaacctact tttcctacca
      301 atcaagaaat tccaagttta attagcggtg ctgatttagt acctattact cctattaccc
      361 cacctttaac tagaacaagc aatagtgcca acaataatgc agcaaatggg atcaatcctc
      421 gctttaaaga cgaagctttt aatgatgttt taatttttga aaatcgccct gcggtttctg
      481 atttttttaac catttttaggc cctagcggag cagctttaac ggtttgggct ttagcacaag
      541 gaaattggat ttggggctat actttaatcg atagcaaagg atttggcgat gctagagttt
      601 ggcaactttt gctttatcct aatgattttg caatgattaa aaatgccaaa accaatactt
      661 gtcttaatgc ttatggtaat ggaattgtcc attatccttg tgatgcaagc aatcacgcac
      721 aaatgtggaa acttatccct atgagcaata cagcggttca aattaaaaat ttaggaaatg
      781 gaaaatgcat acaagcacct attacaaatc tttatggtga ttttcacaag gtttttaaaa
      841 tttttaccgt agagtgtgca aaaaaagata attttgatca acaatggttt ttaactactc
      901 cacctttac cgcaaaacct ttatatcgcc aaggagaggt acgatgaaaa aaattatatg
      961 tttattttta tcttttaacc ttgcttttgc aaatttagaa aattttaatg ttggcacttg
     1021 gaatttgcaa ggctcatccg cagccacaga aagcaaatgg agtgttagtg taagacaact
     1081 tgtaagtgga gcaaacccct tagatatctt aatgatacaa gaagcaggaa ctttaccaag
     1141 aacagccact ccaacaggac gccatgtgca acaaggtgga acacctattg atgaatatga
     1201 gtggaattta ggaactcttt caaggcctga tagggttttt atttattatt ctcgcgttga
     1261 tgtaggagct aatcgtgtaa atttagctat agtttcaaga atgcaagctg aagaagtgat
     1321 tgtttaccct ccacctacta cagtttcaag acccattata ggaattcgca atggaaatga
     1381 tgctttttc aatatccatg ctttagctaa tggaggaaca gatgtaggag caattatcac
     1441 agctgtagat gcacattttg caaatatgcc tcaagttaac tggatgatag caggggattt
     1501 taaccgtgat ccttctacta aacaagtac agtggataga gaattagcaa atagaattag
     1561 agtggttttt ccaactagcg caactcaagc aagcggaggg actcttgatt atgcaattac
     1621 aggaaattca aatagacaac aaacctatac tccaccgctt ttagctgcga ttttaatgct
     1681 tgcaagttta agatctcata tagtttcaga tcattttcca gtaaatttta gaaaattta
     1741 ggacatttaa tatgaaaaaa attattactt tgttttttat gtttataact ttagcctttg
     1801 caactcctac tggagatttg aaagattta ccgaaatggt ttctataaga agcttagaaa
     1861 cgggaatttt tttaagcgcc tttagggata cctcaaaaga tcctattgat caaaattgga
     1921 atattaaaga aattgtttta agcgatgagt taaaacaaaa agataaatta gctgatgaac
     1981 ttccttttgg ttatgtgcaa tttacaaatc caaaagaaag cgatctttgt ttagccatct
     2041 tagaagatgg aacctttgga gcaaaatctt gtcaagatga tctaaaagat ggtaaattag
     2101 aaactgtatt ttctataatg ccaacaacaa cttcagctgt gcaaattcgt tctttagttt
     2161 tggaatctga tgaatgtata gtaacttttt ttaatccaaa tattcctata caaaaacgct
     2221 ttggaatagc ccccttgcacc ctagatccta ttttttttgc tgaagtaaat gaactaatga
     2281 ttataacccc acctttaaca gctgctaccc ctttagaata agattttat cttgttctat
     2341 ttttatattt atttaatatt tatgatatta ctaaaataca caaaataatt aataataata
```

```
2401 caatgtaatt taccttgctc tataattttt ttattttaat gtaatttttt gttacaataa
2461 atttatacat aataattatc ttggaggaaa aattggaaca aattttaaca tggcaacaaa
2521 tttatgaccc tttttcaaat atttggctaa gtgctttagt ggcattttta cctatactat
2581 gttttttagt ttgtttggtt
//
```

SEQ ID 3

```
LOCUS       ECCDTABC                2305 bp    DNA     linear   BCT 29-JAN-1999
DEFINITION  Escherichia coli E6468/62 (O86:H34) cytolethal distending toxin
            (cdtA, cdtB, cdtC) genes, complete cds.
ACCESSION   U03293
VERSION     U03293.1  GI:416213
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 2305)
  AUTHORS   Scott,D.A. and Kaper,J.B.
  TITLE     Cloning and sequencing of the genes encoding Escherichia coli
            cytolethal distending toxin
  JOURNAL   Infect. Immun. 62 (1), 244-251 (1994)
  MEDLINE   94086109
   PUBMED   8262635
REFERENCE   2  (bases 1 to 2305)
  AUTHORS   Scott,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-1993) Scott D.A., University of Maryland School
            of Medicine, Geographic Med., Center for Vaccine Dev., 10 South
            Pine Street, Baltimore, MD 21201, USA
FEATURES             Location/Qualifiers
     source          1..2305
                     /organism="Escherichia coli"
                     /strain="E6468/62 (O86:H34)"
                     /specific_host="Homo sapiens"
                     /db_xref="taxon:562"
                     /clone="pDS7.96"
     gene            201..914
                     /gene="cdtA"
     CDS             201..914
                     /gene="cdtA"
                     /note="GTG start codon"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CdtA"
                     /protein_id="AAD10621.1"
                     /db_xref="GI:416214"
                     /translation="MDKKLIAFLCTLIITGCSNGIGDSPSPPGKNVELVGIPGQGIAV
                     TSNGATPTLGANNTEFPEVSIMSTGGALLTIWARPVRNWLWGYTPFDSVNFGENRNWK
                     VVDGKDAGTVKFVNVAQGTCMEAFKNGVIHNTCDDNSLSQEFQLLPSTNGNVLIRSSA
                     LQTCIRADYLSRTILSPFAFTITLEKCPGAKEETQEMLWAISPPVRAAKPNLIKPELR
                     PFRPLPIPPHDKPDGMEGV"
     gene            911..1732
                     /gene="cdtB"
     CDS             911..1732
                     /gene="cdtB"
                     /codon_start=1
                     /evidence=experimental
                     /transl_table=11
                     /product="CdtB"
                     /protein_id="AAD10622.1"
                     /db_xref="GI:416215"
                     /translation="MKKLLFLLMILPGISFADLSDFKVATWNLQGSNAPTENKWNTHV
                     RQLVTGSGAVDILMVQEAGAVPASATLTEREFSTPGIPMNEYIWNTGTNSRPQELFIY
                     FSRVDAFANRVNLAIVSNRRADEVIVLPPPTVVSRPIIGIRIGNDVFFSTHALANRGV
                     DSGAIVNSVFEFFNRQTDPIRQAANWMIAGDFNRSPATLFSTLEPGIRNHVNIIAPPD
```

```
                           PTQASGGVLDYAVVGNSVSFVLPLLRASLLFGLLRGQIASDHFPVGFIPGRGARR"
        gene               1729..2301
                           /gene="cdtC"
        CDS                1729..2301
                           /gene="cdtC"
                           /codon_start=1
                           /evidence=experimental
                           /transl_table=11
                           /product="CdtC"
                           /protein_id="AAD10623.1"
                           /db_xref="GI:416216"
                           /translation="MKTVIVFFVLLLTGCASEPANQRNLLTQFVGNNAPVDPEPSPVL
                           VNIRNVLTGGIIRNPVGSDFNVNNWVISEVKTNDLDLISAPGGHVQIKNPDGNECFAI
                           LNGQLAVAKQCSESDRNALFTFITSDTGAVQIKSIGSGQCLGNGESITDFRLKKCVDD
                           LGRPFDTVPPGLLWMLNPPLSPAIMSPLTS"
BASE COUNT        641 a      415 c      547 g      702 t
ORIGIN
        1 tattgaatag ttttggggggg aaatataaaga attatatttg agtatgctgt ttgtgactct
       61 gggaataata acgagcatta taataaactt gtttgttttt ctggtttcgc atttcctcat
      121 taatcttgtt ggtaatattt tcgttgcttt gttgtttcta ttttttttata aagaagaggt
      181 ggtgcagagg aggaaataca gtggataaaa aactaattgc attttttgtgc acacttataa
      241 ttactggttg ctcgaatggg atcggtgatt caccttcacc tccgggaaaa aatgtagaat
      301 tggttggaat ccctggacaa ggtattgcag tgacttcaaa cggtgcaact ccaacacttg
      361 gagccaacaa cactgagttt cctgaagttt caataatgag cactggtggg gcgctgctta
      421 ctatttgggc cagacctgtt cgtaactggc tttgggggta tactcctttt gattcagtaa
      481 attttggtga gaatcggaac tggaaggttg tggatgggaa agatgccggc acagtgaaat
      541 ttgttaatgt tgcccagggg acttgcatgg aggcctttaa aaacgggggtg atacataata
      601 cctgtgatga taactcgtta tctcaggagt ttcagttact gccttctact aatggtaatg
      661 tgcttataag aagtagtgcc ttgcagacgt gtataagagc agactattta agcagaacta
      721 tattgtcacc gtttgctttt acaatcaccc ttgagaaatg ccctggtgca aaagaagaaa
      781 cgcaagaaat gctatgggca ataagtccac ctgtcagagc ggcaaaacca aatctgatta
      841 agccagagtt aagaccattc agaccattgc caattccacc tcatgacaaa cctgatggaa
      901 tggagggagt atgaaaaaat tattattcct gttaatgatt ttgccgggta tttctttgc
      961 agatttaagc gatttttaaag ttgcaacctg gaatttgcag ggttcaaatg caccgacaga
     1021 aaataaatgg aacacacatg tccgacaact tgttacggga agtggtgctg ttgatatcct
     1081 gatggttcag gaggcagggg cagtaccagc ttctgcaacg ttgactgagc gagaatttag
     1141 cactcctggt attccgatga atgagtatat ctggaatacc ggaaccaata gtcgcccaca
     1201 ggagttgttt atatatttct cacgtgttga tgcattcgct aacagagtaa atcttgcgat
     1261 tgtttcaaac agaagagctg atgaggtgat tgtattacct cctccaactg ttgtatcacg
     1321 accgatcatc ggcattagaa ttggtaatga tgttttcttc tcaacccatg cattggcgaa
     1381 tcggggcgtg gattcaggag caattgtaaa tagtgttttt gagttcttca acagacaaac
     1441 ggatcctata agacaggccg ctaactggat gattgcagga gattttaacc gttcaccggc
     1501 tacactattt tcaactcttg aaccagggat tcgcaatcat gtaaatatta ttgctccacc
     1561 agatccaacg caagccagtg gtggtgttct tgattatgca gtagttggaa attcagtgag
     1621 ctttgttctt cctctgttga gggcctcgtt gttattcgga ttattaagag ggcaaattgc
     1681 ctctgatcat tttccggttg gctttattcc tggaagagga gcaagaagat gaaaacagtt
     1741 attgtgtttt ttgtttttact gctgacaggt tgtgcttctg aacctgcaaa tcagcgtaat
     1801 cttcttactc agtttgtcgg caacaatgcc cctgtagacc ctgaacccag tccagtattg
     1861 gttaatatca gaaacgttct tacagggggg ataatccgaa atcctgttgg cagtgacttt
     1921 aatgtaaata attgggttat atctgaagta aagactaatg atttggattt gatatcggca
     1981 ccgggagggc atgttcagat taaaaatcct gatggcaatg aatgctttgc tattctaaac
     2041 gggcaattgg cagtggctaa gcagtgctct gaaagtgacc gtaacgcatt gtttacattt
     2101 ataaccagtg atactggggc tgtgcaaatc aagtcaatag gaagcggtca atgcctaggg
     2161 aatggagaga gcattacaga tttcaggtta aaaaaatgtg ttgatgatct gggcgtcct
     2221 tttgatacgg tgccgccggg gttactctgg atgctgaatc caccattatc tccggcaata
     2281 atgtctccat taacgagctg atctg
//
```

SEQ ID 4

```
LOCUS       ECU04208                2600 bp    DNA       linear   BCT 07-JUN-1994
DEFINITION  Escherichia coli 9142-88 cytolethal distending toxin (cdtA, cdtB,
            and cdtC) genes, complete cds.
ACCESSION   U04208
VERSION     U04208.1  GI:436944
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 2600)
  AUTHORS   Pickett,C.L., Cottle,D.L., Pesci,E.C. and Bikah,G.
  TITLE     Cloning, sequencing, and expression of the Escherichia coli
            cytolethal distending toxin genes
  JOURNAL   Infect. Immun. 62 (3), 1046-1051 (1994)
  MEDLINE   94156453
   PUBMED   8112838
REFERENCE   2  (bases 1 to 2600)
  AUTHORS   Pickett,C.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1993) C.L. Pickett, University of Kentucky,
            Microbiology and Immunology, Dept. of Micro and Immuno - Chandler
            Medical Center, Lexington, KY 40536, USA
FEATURES             Location/Qualifiers
     source          1..2600
                     /organism="Escherichia coli"
                     /strain="9142-88"
                     /db_xref="taxon:562"
     gene            408..1184
                     /gene="cdtA"
     CDS             408..1184
                     /gene="cdtA"
                     /codon_start=1
                     /transl_table=11
                     /product="CdtA"
                     /protein_id="AAA18785.1"
                     /db_xref="GI:436945"
                     /translation="MANKRTPIFIAGILIPILLNGCSSGKNKAYLDPKVFPPQVEGGP
                     TVPSPDEPGLPLPGPGPALPTNGAIPIPEPGTAPAVSLMNMDGSVLTMWSRGAGSSLW
                     AYYIGDSNSFGELRNWQIMPGTRPNTIQFRNVDVGTCMTSFPGFKGGVQLSTAPCKFG
                     PERFDFQPMATRNGNYQLKSLSTGLCIRANFLGRTPSSPYATTLTMERCPSSGEKNFE
                     FMWSISEPLRPALATIAKPEIRPFPPQPIEPDEHSTGGEQ"
     gene            1181..1990
                     /gene="cdtB"
     CDS             1181..1990
                     /gene="cdtB"
                     /codon_start=1
                     /transl_table=11
                     /product="CdtB"
                     /protein_id="AAA18786.1"
                     /db_xref="GI:436946"
                     /translation="MKKYIISLIVFLSFYAQADLTDFRVATWNLQGASATTESKWNIN
                     VRQLISGENAVDILAVQEAGSPPSTAVDTGTLIPSPGIPVRELIWNLSTNSRPQQVYI
                     YFSAVDALGGRVNLALVSNRRADEVFVLSPVRQGGRPLLGIRIGNDAFFTAHAIAMRN
                     NDAPALVEEVYNFFRDSRDPVHQALNWMILGDFNREPADLEMNLTVPVRRASEIISPA
                     AATQTSQRTLDYAVAGNSVAFRPSPLQAGIVYGARRTQISSDHFPVGVSRR"
     gene            2005..2550
                     /gene="cdtC"
     CDS             2005..2550
                     /gene="cdtC"
```

```
                    /codon_start=1
                    /transl_table=11
                    /product="CdtC"
                    /protein_id="AAA18787.1"
                    /db_xref="GI:436947"
                    /translation="MKKLAIVFTMLLIAGCSSSQDSANNQIDELGKENNSLFTFRNIQ
               SGLMIHNGLHQHGRETIGWEIVPVKTPEEALVTDQSGWIMIRTPNTDQCLGTPDGRNL
               LKMTCNSTAKKTLFSLIPSTTGAVQIKSVLSGLCFLDSKNSGLSFETGKCIADFKKPF
               EVVPQSHLWMLNPLNTESPII"
BASE COUNT        834 a      555 c      508 g      703 t
ORIGIN
        1 attaacaaat tacaacaaag atcacattaa ataaaaattg acaattgacc tgtagttcat
       61 catttgtaaa ttcatgattt atatcaatca cgctttgtgt tcggagtaag cttataaatt
      121 acaaaaacga ttaaataaaa aaccacacaa taatattaaa taaaaatacg gtatcgactg
      181 cttttgtttc aaaaggaatt gctattaaaa ctatatactt tcatttagtt ttatcaatta
      241 atgcataact tcaaatgtaa catcaaaaac aatacacctc aaaacaatca cacaaagcaa
      301 caaggacacc caaacaacta aggcactaat aaaaaggaga gtcccaatgt aattctttta
      361 ttcttccatt aatttctact atctttatca taataaggac cataataatg gctaacaaac
      421 gtacacctat ttttatagct ggaatcttga tccccatttt attaaatggt tgctcatcag
      481 gaaaaaataa agcttatctt gaccccaaag ttttccctcc tcaagtggaa ggaggaccaa
      541 ccgttccttc ccccgatgag cccggacttc cattgcccgg gccaggaccg gcgctgccca
      601 caaatggcgc aatccctatc cctgaaccag gtaccgcacc cgcagtatct ttaatgaata
      661 tggatggctc agttctaaca atgtggagcc gcggagctgg ttcatcgtta tgggcgtatt
      721 atatcggcga ctccaattca tttggggaac tacgtaattg gcagattatg cccggaacca
      781 ggccaaatac gatacagttt cgcaatgtag acgttggtac ctgtatgaca agtttcccag
      841 gatttaaagg gggagtacaa ctttctacag caccttgcaa gtttggaccg gaacgtttcg
      901 atttccagcc aatggcaaca cgcaatggta attaccagtt aaaatcttta tctacaggtt
      961 tatgcatcag agcgaatttt ttaggaagaa caccatcatc tccgtacgca acgacattaa
     1021 caatggagcg ttgcccatca agtggagaga aaaactttga attcatgtgg tccataagcg
     1081 aaccattaag gcctgctctg gccactattg ccaagccaga aatacgccca tttccaccac
     1141 agccaataga accagatgag cattcaactg gaggagaaca atgaaaaaat atattatatc
     1201 tctgatagtg tttttatcat tttacgctca agcagattta actgatttc gcgttgcgac
     1261 ctggaatctt caaggtgcat ccgctacgac tgaaagtaaa tggaatataa atgtccggca
     1321 attaatttct ggtgaaaatg ctgtagacat tttagctgta caagaggcag gctctccgcc
     1381 gtcaacggct gtagatacag gtacacttat tccttcccca ggaattcccg tccgagagct
     1441 tatctggaac ttgtcgacaa atagcaggcc acagcaagta tatatatatt tttccgctgt
     1501 tgatgccctc ggtggaagag tcaatcttgc tctggttagc aatcggcggg ccgatgaagt
     1561 gtttgttctt agtcctgtaa gacaaggtgg acgaccattg cttggcatac gaattggtaa
     1621 tgatgcattt ttcactgcac acgccatagc tatgcgaaac aatgatgccc cggctcttgt
     1681 tgaggaagtg tataacttct tccgcgacag cagagaccca gtacaccagg cgcttaactg
     1741 gatgattctt ggtgatttca accgtgaacc tgcggattta gagatgaacc ttactgttcc
     1801 cgtaagaagg gcatcagaaa ttatttcacc agcggcggca acacaaacca gccagcgaac
     1861 attagattat gcagtagcag gaaactctgt ggcatttaga ccatctccgc tacaagcggg
     1921 aattgtatat ggagccagga gaactcaaat atcttcagat catttccctg ttggcgtatc
     1981 cagacgataa aagaggctat cataatgaaa aaattagcaa ttgttttac tatgctgcta
     2041 atagctggat gctcttcatc acaggattca gctaacaatc agatagatga attaggaaaa
     2101 gaaaacaatt ctctattcac attccgcaat atccaaagtg gcttaatgat ccataatgga
     2161 ttacatcagc atggccgaga gactattgga tgggaaatag tccctgtgaa aacacctgaa
     2221 gaagcacttg ttaccgatca aagcgggtgg ataatgattc gaacgccaaa cacagaccaa
     2281 tgtttaggga cgcctgatgg aaggaacctg ctaaaaatga cgtgtaattc aacagctaag
     2341 aaaactttgt tttctctcat accgtcaaca acaggggcag tacaaatcaa aagcgttctg
     2401 tctgggcttt gtttcttaga tagtaaaaat agcggattaa gttttgaaac ggggaaatgc
     2461 attgctgact tcaaaaaacc atttgaagtt gtaccacaga gccatttgtg gatgttgaac
     2521 ccattaaata ctgaatcgcc tattatttaa tcccatcatc gcattttgcc gggcacataa
     2581 aaagcattat cataataagt
//
```

SEQ ID 5

```
LOCUS       AF373206                468 bp    DNA     linear   BCT 02-MAY-2002
DEFINITION  Escherichia coli strain MBU. E 412 cytolethal distending toxin B
            (cdtB) gene, partial cds.
ACCESSION   AF373206
VERSION     AF373206.1  GI:20385547
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Bouzari,S., Oloomi,M. and Zarepoor,M.
  TITLE     Identification of cdtB homologs in diarrheagenic E.coli isolates in
            Iran
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 468)
  AUTHORS   Bouzari,S., Oloomi,M. and Zarepoor,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-2001) Molecular Biology, Institute Pasteur of
            Iran, Pasteur Ave., Tehran 13164, Iran
FEATURES             Location/Qualifiers
     source          1..468
                     /organism="Escherichia coli"
                     /strain="MBU. E 412"
                     /db_xref="taxon:562"
     gene            <1..>468
                     /gene="cdtB"
     CDS             <1..>468
                     /gene="cdtB"
                     /note="CdtB"
                     /codon_start=2
                     /transl_table=11
                     /product="cytolethal distending toxin B"
                     /protein_id="AAM21324.1"
                     /db_xref="GI:20385548"
                     /translation="ENKWNTHVRQLVTGSGAVDILMVQEAGAVPASATLTEREFSTPG
                     IPMNEYIWNTGTNSRPQELFIYFSRVDAFANRVNLAIVSNRRADEVIVLPPPTVVSRP
                     IIGIRIGNDVFFSTHALANRGVDSGAIVNSVFEFFNRQTDPIRQAANWMIAGD"
BASE COUNT      130 a     89 c    115 g    134 t
ORIGIN
        1 tgaaaataaa tggaacacac atgtccgaca acttgttacg ggaagtggtg ctgttgatat
       61 cctgatggtt caggaggcag gggcagtacc agcttctgca acgttgactg agcgagaatt
      121 tagcactcct ggtattccga tgaatgagta tatctggaat accggaacca atagtcgccc
      181 acaggagttg tttatatatt tctcacgtgt tgatgcattc gctaacagag taaatcttgc
      241 gattgtttca aacagaagag ctgatgaggt gattgtatta cctcctccaa ctgttgtatc
      301 acgaccgatc atcggcatta gaattggtaa tgatgttttc ttctcaaccc atgcattggc
      361 gaatcggggc gtggattcag gagcaattgt aaatagtgtt tttgagttct tcaacagaca
      421 aacggatcct ataagacagg ccgctaactg gatgattgca ggagattt
//
```

SEQ ID 5

[handwritten marks]

```
LOCUS       G22P1                    2743 bp    mRNA      linear   PRI 05-NOV-2002
DEFINITION  Homo sapiens thyroid autoantigen 70kDa (Ku antigen) (G22P1), mRNA.
ACCESSION   NM_001469
VERSION     NM_001469.2  GI:20070134
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2743)
  AUTHORS   Chan,J.Y., Lerman,M.I., Prabhakar,B.S., Isozaki,O., Santisteban,P.,
            Kuppers,R.C., Oates,E.L., Notkins,A.L. and Kohn,L.D.
  TITLE     Cloning and characterization of a cDNA that encodes a 70-kDa novel
            human thyroid autoantigen
  JOURNAL   J. Biol. Chem. 264 (7), 3651-3654 (1989)
  MEDLINE   89139411
   PUBMED   2917966
REFERENCE   2  (bases 1 to 2743)
  AUTHORS   Reeves,W.H. and Sthoeger,Z.M.
  TITLE     Molecular cloning of cDNA encoding the p70 (Ku) lupus autoantigen
  JOURNAL   J. Biol. Chem. 264 (9), 5047-5052 (1989)
  MEDLINE   89174787
   PUBMED   2466842
REFERENCE   3  (bases 1 to 2743)
  AUTHORS   Griffith,A.J., Craft,J., Evans,J., Mimori,T. and Hardin,J.A.
  TITLE     Nucleotide sequence and genomic structure analyses of the p70
            subunit of the human Ku autoantigen: evidence for a family of genes
            encoding Ku (p70)-related polypeptides
  JOURNAL   Mol. Biol. Rep. 16 (2), 91-97 (1992)
  MEDLINE   92301477
   PUBMED   1608402
REFERENCE   4  (bases 1 to 2743)
  AUTHORS   Tuteja,N., Tuteja,R., Ochem,A., Taneja,P., Huang,N.W.,
            Simoncsits,A., Susic,S., Rahman,K., Marusic,L., Chen,J. et al.
  TITLE     Human DNA helicase II: a novel DNA unwinding enzyme identified as
            the Ku autoantigen
  JOURNAL   EMBO J. 13 (20), 4991-5001 (1994)
  MEDLINE   95045391
   PUBMED   7957065
REFERENCE   5  (bases 1 to 2743)
  AUTHORS   Koike,M., Matsuda,Y., Mimori,T., Harada,Y.N., Shiomi,N. and
            Shiomi,T.
  TITLE     Chromosomal localization of the mouse and rat DNA double-strand
            break repair genes Ku p70 and Ku p80/XRCC5 and their mRNA
            expression in various mouse tissues
  JOURNAL   Genomics 38 (1), 38-44 (1996)
  MEDLINE   97124844
   PUBMED   8954777
REFERENCE   6  (bases 1 to 2743)
  AUTHORS   Baumann,P. and West,S.C.
  TITLE     DNA end-joining catalyzed by human cell-free extracts
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (24), 14066-14070 (1998)
  MEDLINE   99045638
   PUBMED   9826654
REFERENCE   7  (bases 1 to 2743)
  AUTHORS   Goedecke,W., Eijpe,M., Offenberg,H.H., van Aalderen,M. and
            Heyting,C.
  TITLE     Mre11 and Ku70 interact in somatic cells, but are differentially
            expressed in early meiosis
  JOURNAL   Nat. Genet. 23 (2), 194-198 (1999)
```

```
     MEDLINE    99438394
      PUBMED    10508516
     COMMENT    PROVISIONAL REFSEQ: This record has not yet been subject to final
                NCBI review. The reference sequence was derived from BC008343.1.
                On Apr 8, 2002 this sequence version replaced gi:4503840.
                Summary: The G22P1 gene encodes subunit p70 of the p70/p80
                autoantigen. The p70/p80 autoantigen consists of 2 proteins of
                molecular mass of approximately 70,000 and 80,000 daltons that
                dimerize to form a 10 S DNA-binding complex. See MIM 194364 for
                discussion of the gene encoding the p80 subunit. Exchange of
                immunologic reagents showed that the p70/p80 autoantigen is
                identical to the Ku antigen, the Ki antigen, and the 86- to 70-kD
                protein complex. The p70/p80 complex binds to the ends of
                double-stranded DNA in a cell cycle-dependent manner, being
                associated with chromosomes of interphase cells, followed by
                complete dissociation from the condensing chromosomes in early
                prophase. Both p70 and p80 contain phosphoserine residues. A role
                for the antigen in DNA repair or transposition has been
                proposed.[supplied by OMIM].
    FEATURES             Location/Qualifiers
         source          1..2743
                         /organism="Homo sapiens"
                         /db_xref="taxon:9606"
                         /chromosome="22"
                         /map="22q13.2-q13.31"
                         /clone="MGC:15860 IMAGE:3509801"
                         /tissue_type="Brain, neuroblastoma"
                         /clone_lib="NIH_MGC_19"
                         /lab_host="DH10B-R"
                         /note="Vector: pOTB7"
         gene            1..2743
                         /gene="G22P1"
                         /note="synonyms: ML8, KU70, TLAA, D22S671, D22S731"
                         /db_xref="LocusID:2547"
                         /db_xref="MIM:152690"
         CDS             656..2485
                         /gene="G22P1"
                         /note="thyroid autoantigen 70kD (Ku antigen)"
                         /codon_start=1
                         /product="thyroid autoantigen 70kDa (Ku antigen)"
                         /protein_id="NP_001460.1"
                         /db_xref="GI:4503841"
                         /db_xref="LocusID:2547"
                         /db_xref="MIM:152690"
                         /translation="MSGWESYYKTEGDEEAEEEQEENLEASGDYKYSGRDSLIFLVDA
                         SKAMFESQSEDELTPFDMSIQCIQSVYISKIISSDRDLLAVVFYGTEKDKNSVNFKNI
                         YVLQELDNPGAKRILELDQFKGQQGQKRFQDMMGHGSDYSLSEVLWVCANLFSDVQFK
                         MSHKRIMLFTNEDNPHGNDSAKASRARTKAGDLRDTGIFLDLMHLKKPGGFDISLFYR
                         DIISIAEDEDLRVHFEESSKLEDLLRKVRAKETRKRALSRLKLKLNKDIVISVGIYNL
                         VQKALKPPPIKLYRETNEPVKTKTRTFNTSTGGLLLPSDTKRSQIYGSRQIILEKEET
                         EELKRFDDPGLMLMGFKPLVLLKKHHYLRPSLFVYPEESLVIGSSTLFSALLIKCLEK
                         EVAALCRYTPRRNIPPYFVALVPQEEELDDQKIQVTPPGFQLVFLPFADDKRKMPFTE
                         KIMATPEQVGKMKAIVEKLRFTYRSDSFENPVLQQHFRNLEALALDLMEPEQAVDLTL
                         PKVEAMNKRLGSLVDEFKELVYPPDYNPEGKVTKRKHDNEGSGSKRPKVEYSEEELKT
                         HISKGTLGKFTVPMLKEACRAYGLKSGLKKQELLEALTKHFQD"
         misc_feature    764..2314
                         /gene="G22P1"
                         /note="ku; Region: A family of DNA binding domains
                         including those in Ku70/80"
                         /db_xref="CDD:LOAD_ku"
```

```
misc_feature     764..1408
                 /gene="G22P1"
                 /note="Ku_N; Region: Ku70/Ku80 N-terminal alpha/beta
                 domain. The Ku heterodimer (composed of Ku70 and Ku80)
                 contributes to genomic integrity through its ability to
                 bind DNA double-strand breaks and facilitate repair by the
                 non-homologous end-joining pathway. This is the amino
                 terminal alpha/beta domain. This domain only makes a small
                 contribution to the dimer interface. The domain comprises
                 a six stranded beta sheet of the Rossman fold"
                 /db_xref="CDD:pfam03731"
misc_feature     1451..2059
                 /gene="G22P1"
                 /note="ku; Region: Ku70/Ku80 beta-barrel domain. The Ku
                 heterodimer (composed of Ku70 and Ku80) contributes to
                 genomic integrity through its ability to bind DNA
                 double-strand breaks and facilitate repair by the
                 non-homologous end-joining pathway. This is the central
                 DNA-binding beta-barrel domain. This domain is found in
                 both the Ku70 and Ku80 proteins that form a DNA binding
                 heterodimer"
                 /db_xref="CDD:pfam02735"
misc_feature     1577..2017
                 /gene="G22P1"
                 /note="Ku78; Region: Ku70 and Ku80 are 70kDa and 80kDa
                 subunits of the Lupus Ku autoantigen"
                 /db_xref="CDD:smart00559"
misc_feature     2066..2332
                 /gene="G22P1"
                 /note="Ku_C; Region: Ku70/Ku80 C-terminal arm. The Ku
                 heterodimer (composed of Ku70 and Ku80) contributes to
                 genomic integrity through its ability to bind DNA
                 double-strand breaks and facilitate repair by the
                 non-homologous end-joining pathway. This is the C terminal
                 arm. This alpha helical region embraces the beta-barrel
                 domain pfam02735 of the opposite subunit"
                 /db_xref="CDD:pfam03730"
misc_feature     2372..2476
                 /gene="G22P1"
                 /note="SAP; Region: SAP domain. The SAP (after SAF-A/B,
                 Acinus and PIAS) motif is a putative DNA binding domain
                 found in diverse nuclear proteins"
                 /db_xref="CDD:pfam02037"
misc_feature     2372..2473
                 /gene="G22P1"
                 /note="SAP; Region: Putative DNA-binding (bihelical) motif
                 predicted to be involved in chromosomal organisation"
                 /db_xref="CDD:smart00513"
variation        1472
                 /gene="G22P1"
                 /allele="A"
                 /allele="G"
                 /db_xref="dbSNP:2229034"
variation        2256
                 /gene="G22P1"
                 /note="WARNING: map location ambiguous"
                 /allele="A"
                 /allele="T"
                 /db_xref="dbSNP:1803107"
variation        2402
```

```
                         /gene="G22P1"
                         /allele="A"
                         /allele="G"
                         /db_xref="dbSNP:3179615"
            variation    2434
                         /gene="G22P1"
                         /allele="G"
                         /allele="T"
                         /db_xref="dbSNP:132788"
BASE COUNT      732 a     629 c     743 g     639 t
ORIGIN
        1 gcgggccgtt atccatttgt gttgttcgcc agctaggcct ggcctcgtcc cgcttcgctc
       61 ggtcggtctc gcgcgccccc atagccttgc tagagggtta gcgttagcct taagtgtgcg
      121 aatccgagga gcagcgacag actcgagacc acgctccttc ctcgggaagg aggcggcacc
      181 tcgcgtttga ggcccgcctg cgtttgaggc ccgcctgcgc ttgcggcccg cctgcgcttg
      241 aggcctgtct gcgtttgaga tctcattggg cgtgattgag gaatttgggg aggtttttgg
      301 gcggtattga ggacgagggg gtccgttagt cagcatagaa tcctggagcg ggaatccctc
      361 accgtctaaa tggcgtcggg ggcgggacct ccgggatctg gcttccgcgg gccgccgccg
      421 gccctgaaac gtgagggata gctgagatga ggcagctact gggatggccc ccatgcgcat
      481 ttacatgcag tccgactgcc gagctttcga ggcagcagga tttaccgtcc acattcctca
      541 ctactaacca agcttttaga acagatctca caagaaccta gaggtcggta tttttttcgat
      601 ttaaatttgc ctgttactga cgttaacgtc tttcgcctag tgagcagtag ccaacatgtc
      661 agggtgggag tcatattaca aaaccgaggg cgatgaagaa gcagaggaag aacaagaaga
      721 gaaccttgaa gcaagtggag actataaata ttcaggaaga gatagtttga ttttttttggt
      781 tgatgcctcc aaggctatgt ttgaatctca gagtgaagat gagttgacac cttttgacat
      841 gagcatccag tgtatccaaa gtgtgtacat cagtaagatc ataagcagtg atcgagatct
      901 cttggctgtg gtgttctatg gtaccgagaa agacaaaaat tcagtgaatt ttaaaaatat
      961 ttacgtctta caggagctgg ataatccagg tgcaaaacga attctagagc ttgaccagtt
     1021 taaggggcag cagggacaaa aacgtttcca agacatgatg ggccacggat ctgactactc
     1081 actcagtgaa gtgctgtggg tctgtgccaa cctctttagt gatgtccaat tcaagatgag
     1141 tcataagagg atcatgctgt tcaccaatga agacaacccc catggcaatg acagtgccaa
     1201 agccagccgg gccaggacca aagccggtga tctccgagat acaggcatct tccttgactt
     1261 gatgcacctg aagaaacctg ggggctttga catatccttg ttctacagag atatcatcag
     1321 catagcagag gatgaggacc tcaggggttca cttttgaggaa tccagcaagc tagaagacct
     1381 gttgcggaag gttcgcgcca aggagaccag gaagcgagca ctcagcaggt taaagctgaa
     1441 gctcaacaaa gatatagtga tctctgtggg catttataat ctggtccaga aggctctcaa
     1501 gcctcctcca ataaagctct atcgggaaac aaatgaacca gtgaaaacca agacccggac
     1561 ctttaataca agtacaggcg gtttgcttct gcctagcgat accaagaggt ctcagatcta
     1621 tgggagtcgt cagattatac tggagaaaga ggaaacagaa gagctaaaac ggtttgatga
     1681 tccaggtttg atgctcatgg gtttcaagcc gttggtactg ctgaagaaac accattacct
     1741 gaggccctcc ctgttcgtgt acccagagga gtcgctggtg attgggagct caaccctgtt
     1801 cagtgctctg ctcatcaagt gtctggagaa ggaggttgca gcattgtgca gatacacacc
     1861 ccgcaggaac atccctcctt attttgtggc tttggtgcca caggaagaag agttggatga
     1921 ccagaaaatt caggtgactc ctccaggctt ccagctggtc tttttaccct ttgctgatga
     1981 taaaaggaag atgcccttta ctgaaaaaat catggcaact ccagagcagg tgggcaagat
     2041 gaaggctatc gttgagaagc ttcgcttcac atacagaagt gacagctttg agaaccccgt
     2101 gctgcagcag cacttcagga acctggaggc cttggccttg gatttgatgg agccggaaca
     2161 agcagtggac ctgacattgc ccaaggttga agcaatgaat aaaagactgg ctccttggt
     2221 ggatgagttt aaggagcttg tttacccacc agattacaat cctgaaggga aagttaccaa
     2281 gagaaaacac gataatgaag gttctggaag caaaaggccc aaggtggagt attcagaaga
     2341 ggagctgaag acccacatca gcaagggtac gctgggcaag ttcactgtgc ccatgctgaa
     2401 agaggcctgc cgggcttacg ggctgaagag tggtctgaag aagcaggagc tgctggaagc
     2461 cctcaccaag cacttccagg actgaccaga ggccgcgcgt ccagctgccc ttccgcagtg
     2521 tggccaggct gcctggcctt gtcctcagcc agttaaaatg tgtttctcct gagctaggaa
     2581 gagtctaccc gacataagtc gagggacttt atgttttga ggctttctgt tgccatggtg
     2641 atggtgtagc cctcccactt tgctgttctt tactttactg cctgaataaa gagccctaag
     2701 tttgtactaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaa aaa
//
```

SEQ ID 7

```
LOCUS        AF093759              2493 bp    DNA      linear    PRI 14-AUG-2001
DEFINITION   Homo sapiens heat shock protein HSP70 (HSPA7) gene, complete
             sequence.
ACCESSION    AF093759
VERSION      AF093759.1  GI:4139180
KEYWORDS     .
SOURCE       Homo sapiens
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 2493)
  AUTHORS    Parsian,A.J., Sheren,J.E., Tao,T.Y., Goswami,P.C., Malyapa,R., Van
             Rheeden,R., Watson,M.S. and Hunt,C.R.
  TITLE      The human Hsp70B gene at the HSPA7 locus of chromosome 1 is
             transcribed but non-functional
  JOURNAL    Biochim. Biophys. Acta 1494 (1-2), 201-205 (2000)
  MEDLINE    20525459
   PUBMED    11072087
REFERENCE    2  (bases 1 to 2493)
  AUTHORS    Hunt,C.R., Malyapa,R., Parsian,A.J., Goswami,P.C., Van Rheeden,R.
             and Watson,M.S.
  TITLE      Direct Submission
  JOURNAL    Submitted (22-SEP-1998) Radiation Oncology, Washington University
             School of Medicine, 4511 Forest Park Blvd., St. Louis, MO 63108,
             USA
FEATURES             Location/Qualifiers
     source          1..2493
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1q21.3"
     gene            233..2149
                     /gene="HSPA7"
                     /note="heat-inducible heat shock protein HSP70; possible
                     pseudogene"
BASE COUNT      554 a    676 c    794 g    469 t
ORIGIN
        1 cccgggcggg cgggcgggag gctctcgact gggcgggaag gtgcgggaag gttcgcggcg
       61 gcggggtcgg ggaggtgcaa aaggatgaaa agcccgtgga agcggagctg agcagatccg
      121 agccgggctg gcggcagaga aaccgcaggg agagcctcac tgctgagcgc ccctcgacgg
      181 cggagcggca gcagcctccg tggcctccag catccgacaa gaagcttcag ccatgcaggc
      241 cccacgggag ctcgcggtgg gcatcgacct gggcaccacc tactcgtgcg tgggcgtgtt
      301 tcagcagggc cgcgtggaga tcctggccaa cgaccagggc aaccgcacca cgcccagcta
      361 cgtggccttc accgacaccg agcggctggt cggggacgcg ccaagagcc aggcggccct
      421 gaacccccac aacaccgtgt cgatgccaa gcggctgatc gggcgcaagt tcgcggacac
      481 cacggtgcag tcggacatga agcactggcc cttccaggtg gtgagcgagg gcggcaagcc
      541 caaggtgcgc gtatgctacc gcggggagga caagacgttc taccccgagg agatctcgtc
      601 catggtgctg agcaagatga aggagacggc cgaggcgtac ctgggccagc ccgtgaagca
      661 cgcagtgatc accgtgccca cctatttcag taactcgcag cgccaggcca ccaaggacgc
      721 gggggccatc gcggggctca aggtgctgcc gatcatcaat gaggccacgg cagcagccat
      781 cgcctatggg ctggaccggc ggggcgcggg aaagcgcaac gtgctcattt ttgacctggg
      841 tgggggcacc ttcgatgtgt cggttctctc cattgacgcc ggtgtctttg aggtgaaagc
      901 cactgctgga gatacccacc tgggaggaga ggacttcgac aaccggctcg tgaaccactt
      961 catggaagaa ttccggcgga agcatgggaa ggacctgagc gggaacaagc gtgccctgcg
     1021 caggctgcgc acagcctgtg agcgcgccaa gcgcacccg tcctccagca cccaggccac
     1081 cctggagata gactccctgt tcgagggcgt ggacttctac aagtccatca ctcgtgcccg
     1141 ctttgaggaa ctgtgctcag acctcttccg cagcaccctg gagccggtgg agaaggccct
     1201 gcgggatgcc aagctggaca aggcccagat tcatgacttc gtcctggggg gagggctcca
     1261 ctcgcatccc caaggtgcag aagttgctgc aggacttctt caacggcaag gagctgaaca
     1321 agagcatcaa ccctgatgag gctgtggcct atgggtctgc tgtgcaggcg ccgtgttga
```

```
1381 tggggggacaa atgtgagaaa gtgcaggatc tcctgctgct ggatgtggct cccctgtctc
1441 tggggctgga gacagcaggt ggggtgatga ccacgctgat ccagaggaac gccactatcc
1501 ccaccaagca gacccagact ttcaccacct actcggacaa ccagcctggg gtcttcatcc
1561 aggtgtatga ggttgagagg gccatgacca aggacaacaa cctgctgggg cgttttgaac
1621 tcattggcat ccctcctgcc ccacatggag tcccccagat agaggtgacg tttgacattg
1681 atgctaatgg catcctgagc gtgacagcca ctgacaggag cacaggtaag ctaacaaga
1741 tcaccaatga caagggccgg ctgagcaagg aggaggtgga gaggatggtt catgaagccg
1801 agcagtacgg ggctgaggat gaggcccaga gggacagagt ggctgccaaa aactcgctgg
1861 aggcccatgt cttccatgtg aaaggttctt tgcaagagga aagccttagg dacaagattc
1921 ccgaagagga caggcgcaaa gtgcaagaca agtgtcagga agtccttgcc tggctggagc
1981 acaaccagct ggcagagaag gaggagtatg agcatcagaa gagggagctg gagcaaatct
2041 gtcgccccat cttctccagg ctctatgggg ggcctggtgt ccctgggggc agcagttgta
2101 gcgctcaagc ccaccagggg gaccccagca ccggccccat cattgaggag gttgattgaa
2161 tggcccttcg tgataagtca gctgtgactg tcagggctat gctatgggcc ttctagactg
2221 tcttctatga tcctgccctt cagagatgaa gggcttgggg gggtcttccc tccaaagcta
2281 gaactttctt tccaggataa ctgaagtctt ttgactttt ggggggaggg cggttcatcc
2341 tcttctgctt caaataaaaa gtcattaatt tattaaaact tgtgtggcac tttaacattg
2401 ctttcaccta tattttgtgt attttgttac ttgtatgtat gaattttgtt atgtaaaata
2461 tagttataga cctaaataaa cttttaaaac tcc
//
```